Assessment may conceal therapeutic benefit: findings from a randomized controlled trial for hazardous drinking

Kypros Kypri, PhD*^{1, 2}
John D. Langle, PhD²
John B. Saunders, MB MD³
Martine L. Cashell-Smith, BCom²

*Corresponding author:

¹School of Medical Practice and Population Health, University of Newcastle,
2 Edison St, Adamstown Heights NSW Australia
Tel: +61 2 4927 5957, Fax: +61 2 4924 6208, kypri@tpg.com.au

²Injury Prevention Research Unit, Department of Preventive and Social Medicine,
University of Otago, Dunedin, New Zealand

³Centre for Drug and Alcohol Studies, Department of Psychiatry, School of Medicine,
University of Queensland, Brisbane, Australia

Word count

Abstract: 259

Text: 3871

Abstract

Context: The concept that assessment of a person's health status without subsequent intervention has beneficial effects in itself has stimulated much interest in underlying psychological mechanisms, methodological implications, and its public health potential. There have, however, been few experimental studies of assessment effects.

Aim: To test the hypothesis that assessment in itself produces a reduction in hazardous drinking.

Design and Setting: Two conditions (A: leaflet only and B: leaflet and assessment but no intervention) of a four-arm randomized controlled trial with enrolment in March-April 2003.

Participants: 975 students (17-29 years) attending a primary healthcare clinic completed a web-based AUDIT questionnaire. Of 599 who scored ≥ 8 , 576 consented to follow-up and were included in the full four-arm trial, of whom 293 (153 women) were assigned to groups A and B.

Intervention: Group A received an information leaflet at baseline. Group B received the information leaflet and 10 min of web-based assessment 4 weeks later.

Measurements: Drinking frequency, typical quantity, heavy episode frequency, personal problems, and academic problems.

Findings: Baseline mean AUDIT scores were 15.0 (SD=5.4) and 14.9 (SD=5.0) in A and B respectively. Twelve months after baseline, relative to group A, group B reported lower overall consumption (geometric means ratio 0.82, 95% CI: 0.68 to 0.98), fewer heavy drinking episodes (0.66, 0.47 to 0.91), fewer problems (0.81, 0.67 to 0.99), and lower AUDIT scores (beta= -1.63, -0.62 to -2.65).

Conclusions: Brief assessment appeared to reduce hazardous drinking. Controlled trials that rely on assessment may therefore under-estimate treatment effects. Limitations include the possibility of measurement artefact due to social desirability bias.

Assessment is a foundation of all clinical practice for the purpose of diagnosis, treatment planning, and monitoring outcomes. The concept that assessment of a person's health status without subsequent intervention has beneficial effects in itself, has stimulated much interest in underlying psychological mechanisms, its public health implications, and methodological problems (for a review see [1]). The methodological problems relate primarily to the blurring of experimental contrast when assessment has a therapeutic benefit, a form of *contamination* in clinical trials parlance [2]. Even small assessment effects can obscure the effects of brief interventions, whose chief virtue lies in the aggregation of modest therapeutic benefits across populations.

Perhaps the most famous assessment effect was that noticed more than 70 years ago by Professor Elton Mayo in the Hawthorne factory of the American company Western Electric. Mayo was studying the effects of environmental conditions, such as lighting and temperature, on workers' behaviour. He observed that workers increased their productivity not as a result of changes in work conditions but as a result of management taking special interest in such improvements [3]. The eponymous *Hawthorne Effect* refers to the phenomenon of individuals' behaviour changing because they know they were being studied.

Assessment effects (also called 'assessment reactivity' or 'subject reactivity') are of clear importance in the drug and alcohol treatment field, which relies on the assessment of patients' substance use and related problems for the purpose of planning and delivering treatment. They are also important for research, in study design, data analysis, and interpretation of results. The possibility that assessment produces therapeutic benefits has

been the subject of much scientific discussion [4, 5]. In their critical analysis of assessment effect studies, Clifford and Maisto wrote:

“The disentangling of research and treatment effects is critical if research is to improve decisions made about alcohol treatment. The separation of subject reactivity and treatment effects is essential, to the degree that it is present in the typical randomized controlled trial (RCT), because most people presenting for alcohol treatment are never exposed to the research component.” (p787) [1].

In a pilot RCT of electronic screening and brief intervention at a New Zealand university student health service, reductions of 20-30% in hazardous alcohol consumption lasting six months were reported, with some attenuation of differences between the control and intervention groups from the 6-week follow-up assessment to 6-month follow-up assessment [6]. A process evaluation which followed the pilot trial suggested an explanation for this attenuation. Several control group participants said that the 6-week assessment had a moderating effect on them. It is possible that the reduction in effect sizes seen at six months resulted from the partial exposure of the control group to the intervention (in the form of a web-based assessment) at six weeks, i.e., an assessment effect.

Furthermore, an analysis of written comments during the trial, suggesting that completion of a drinking diary brought excessive consumption to participants' attention, supports the notion of an assessment effect [7]. The presence of assessment effects has been postulated in other brief intervention trials, in which consumption is often found to trend downwards for both intervention and control groups subsequent to initial assessment [8, 9]. Some of this reduction may be a result of regression to the mean [10], but an assessment effect remains a viable explanation for at least part of this phenomenon. This hypothetical effect may feature in other alcohol and drug treatment trials, and despite often being discussed as a potential explanation for declining consumption in those trials,

there have been only a few published studies in the addiction literature seeking to address the issue experimentally. According to Clifford and Maisto, those studies have been "...hampered by such methodological shortcomings as insufficient power, failure to randomly assign subjects and treatment fidelity problems." (p 788) [1].

We were interested in the possibility of assessment alone having a therapeutic effect for two reasons: (1) because of the potential for brief assessment to be used for public health benefit, and (2) because the use of baseline assessment in brief intervention trials might be concealing therapeutic effects. As part of a four-arm randomised controlled trial of electronic screening and brief intervention [6], we included two control groups: one received only a leaflet on the effects of alcohol, while a second received a leaflet and an assessment, but no intervention. The aim of this aspect of the trial was to test the hypothesis that assessment in itself would produce a reduction in hazardous drinking at 6- and 12-month follow-up.

Methods

Study design

The overall study was a four-arm double-blind randomized controlled trial. Students who screened positive for hazardous drinking were assigned to one of two groups receiving web-based assessment and personalized feedback on their drinking (brief intervention), or were assigned to one of two control groups, which received a leaflet on the health effects of alcohol, with or without assessment four weeks later. Only the latter two arms are examined here (Figure 1). Ethical approval was granted by the University of Otago Ethics Committee.

<Figure 1>

Sample size estimates

Sample size estimates for the trial were based on an effect size of 0.37, the mean observed in the pilot trial [6]. For power of 0.80 with alpha of 0.05 (two-sided), 114 individuals per group were required for analysis. Making allowance for 20% attrition at 12 months, 143 ($114/0.80$) individuals per group had to be recruited at baseline.

Sampling of service users

We sought to maximize the generalizability of the trial results by obtaining a representative sample of service users and encouraging a high level of participation. We opted for a selection protocol which would minimize the risk of systematic biases, and allow for measurement of the potential bias resulting from self selection. Each week (Monday-Friday) of the sampling period was broken into 10 sessions: five morning sessions: 9am-12.30pm, and five afternoon sessions: 1.30-5pm. Based on the ratio of men to women using the service measured in the pilot study [6], we randomly selected two of the 10 sessions in each week for recruitment of men only, to ensure approximately equal numbers of men and women in the study.

Research assistants were trained in the application of the study protocol, which stipulated that the assistant should invite the next patient leaving the reception desk to participate, go through the informed consent procedure, log the participant into a computer, and return to the reception desk to recruit the next patient. Instances in which a patient appeared too sick or injured or whose command of English was insufficient to participate, were recorded, as were refusals to participate.

The research assistants were informed that covert compliance checks would be conducted on a random basis by a principal investigator (KK). These occurred on two occasions in each of the three weeks of enrolment (total of six checks), by observing assistants for 15 minutes from a concealed vantage point in the service waiting area. Observed compliance with the protocol was 100%.

Blinding and consent

Research staff involved in the trial were not informed of participants' group allocation during intervention or at follow-up. The generation of the allocation sequence and loading of it into the server database were conducted by staff who were not involved in the implementation of the trial on site. That is, they had no contact with study participants.

A two stage recruitment procedure was used, whereby patients were first invited to complete a computerized survey (stage 1: screening). Patients eligible for the study on the basis of screening were asked for consent to be contacted for follow-up surveys (stage 2: assessment and intervention). In accordance with ethical approval, the study was presented to potential participants as a series of surveys on alcohol use, not as a randomized trial. Randomization was effected by computer upon completion of screening.

Participants assigned to the control conditions were presented with a web page thanking them for participating. The research assistant gave each participant an *Alcohol Facts and Effects* leaflet [11] and then initiated the second stage of the consent process, by asking

for contact details to be provided for the purpose of follow-up surveys. Upon completion, the participant was thanked and he or she resumed the wait to be seen by medical staff.

It should be noted that the second stage consent procedure was conducted after participants had completed the web pages used in each of the experimental conditions. Students were asked whether they were willing to complete web-based follow-up assessments up to 12 months later, and to provide contact details for this purpose. If they did not provide contact details they were considered to have declined consent, and they were not included in the trial. We used their baseline data (gender, age, ethnicity, and AUDIT score [12]) to quantify differences between consenters and non-consenters.

The web pages were presented as a seamless series for both groups, and the study had been introduced to participants as a series of surveys of their alcohol use, not an intervention trial. Thus it was intended that participants would not realise they had been assigned to one of four groups, or that their experience of the survey was different to that of other participants. We were confident that this was the case since our pilot work indicated that trial participants were not aware of the true nature of the study [6, 7]. Given that the procedures in the present trial were identical to those in the pilot trial, we are confident in the blinding of participants. Research assistants were not informed of the randomisation sequence and could not have distinguished between persons assigned to either of the control conditions which were treated identically at baseline.

Screening and randomization

Screening consisted of the AUDIT followed by a question about the maximum number of drinks they had consumed in the previous four weeks, and the duration (in hours) of that

episode. Participants scoring 8 or more on the AUDIT and who had consumed more than the recommended upper limit for episodic drinking (4 for women, 6 for men), were randomly assigned by computer to receive either (A) a leaflet only at baseline or (B) a leaflet at baseline and then a 10-minute web-based assessment of their drinking four weeks later. The assessment included components identical to those in the pilot trial [6]: a 14-day retrospective drinking diary [13], the Alcohol Problems Scale (APS) [14], the Academic Role Expectations and Alcohol Scale (AREAS) [14], and questions on perceived drinking norms [15]. The assessment instrument can be viewed at <http://ipru.otago.ac.nz/eSBI2003Demo/Index.html>.

The leaflet, produced by the Alcohol Advisory Council of New Zealand, consisted of four double-sided A5 pages, with information on blood alcohol limits for drink-driving, costs to society of alcohol-related harm, standard drink definitions, typical alcohol content of common beverages, tips for safer drinking, and phone numbers for counselling.

Follow-up assessment

Participants were invited by letter to complete 6- and 12-month follow-up surveys by clicking on a hyperlink sent to their e-mail address. Embedded in the hyperlink was a unique identifier which gained the participant access to their record via a web interface. Included with each letter was a lunch voucher valued at NZ\$4.95, as a token of appreciation for their participation. Reminder e-mails were sent to participants who did not respond, followed by a reminder telephone call. The 6- and 12-month follow-up phases were completed by 24/10/2003, and 25/6/2004 respectively.

Outcome measures

Seven outcome measures were selected:

1. *Frequency of drinking*: number of drinking days in the preceding 2 weeks;
2. *Typical occasion quantity*: number of standard drinks consumed per typical drinking occasion in the preceding 4 weeks;
3. *Total volume*: number of standard drinks consumed in the preceding 2 weeks;
4. *Frequency of very heavy episodes*: number of occasions in the preceding 2 weeks where a threshold of 80/120g ethanol was breached, for women/men respectively;
5. *Personal, social, sexual, and legal consequences of episodic heavy drinking*: number of items endorsed on the APS (range 0-14);[14]
6. *Consequences related to academic performance*: a score on the AREAS (range 0-35).[14]
7. AUDIT score measured at 12 months.

Outcomes 1, 3 and 4 were measured with a retrospective diary for the preceding 14 days.

Outcome 2 was measured with the question: "How many drinks containing alcohol did you have on a typical day when you were drinking in the last 4 weeks?".

Analysis

Given that most of the outcomes were based on counts (e.g., number of days drinking in the last two weeks), the data are not, strictly speaking, continuous. Analysis should therefore be based on distributions for discrete data, such as the Poisson distribution, which has one parameter, meaning that the variance is equal to the mean. An alternative is the negative binomial distribution, which has extra parameters, therefore allowing the variance to be greater than the mean, a feature described as over-dispersion [16]. The

variance was greater than the mean for almost all parameter estimates, indicating that the data are distributed according to the negative binomial distribution. Outcomes 1-5 were analysed using negative binomial regression for panel data. For outcome 6, a continuous scale, we used linear regression analysis for panel data, after log transformation. For outcome 7, we used linear regression analysis for panel data, without log transformation. The models were adjusted for baseline differences by including the baseline AUDIT score. They also included terms for the group, follow-up assessment, and their interaction, using the *xtnbreg* procedure in STATA [17]. The interaction term tested for differences in the assessment effect between the 6- and 12-month follow-up assessments. Because negative binomial regression involves a log link, the results for outcomes 1-6 are presented as ratios (with 95% CI) rather than as differences. Outcome 7 is presented as a difference (the beta coefficient) on the AUDIT scale. The results are presented as the ratio of the geometric mean for the assessment group (B) to that of the leaflet only group (A) at the 6- and 12-month assessments [17], where the geometric mean is the exponent of the arithmetic mean of the log transformed values.

Results

Of 1324 patients checked for eligibility, 1120 were invited to participate and 975 completed screening (87%). Of these, 599 (61%) screened positive for hazardous drinking on the AUDIT (score ≥ 8). Twenty-three declined further involvement (second stage consent), leaving 576 individuals in the full four-arm trial, including 293 patients (153 women) in the two control arms studied here. The median completion time for the two control groups (i.e. for screening alone) was 3.3 minutes. Table 1 presents summary data for the two groups.

< Table 1 >

Participant flow and follow-up

The flow of participants through the trial is illustrated in Figure 1. At 6 months, data were obtained from 124 participants in group A and 122 participants in group B. At 12 months, data were obtained from 126 participants in each of the groups. Baseline AUDIT scores did not differ significantly for those who did and did not complete the follow-up assessments at six months (mean difference = -0.26 points; 95% CI: -1.64 to 1.11) but baseline scores for those who did not complete the 12 month assessment were 2.13 points lower than in those who did complete it (95% CI: -3.64, -0.63). The proportion of participants lost to follow-up did not differ significantly by group at six months (0.02; -0.06, 0.10), or 12 months (0.01; -0.07, 0.09).

Table 2 presents the median and range for the outcomes at 6-month and 12-month follow-up assessments.

< Table 2 >

Assessment effect

Estimates of the assessment effect are presented in Table 3, as the ratio of the geometric mean of the assessment group (B) to that of the leaflet only control group (A). At 6 months, relative to the leaflet only group, the assessment group showed a trend to lower consumption and fewer alcohol-related problems but differences were non-significant. At 12 months, differences for the following outcomes were significant: total consumption (18% difference), heavy episodic drinking (34% difference), personal problems (19% difference), and AUDIT score (1.6 points difference). Analysis of the results by gender,

which is reported elsewhere [18], showed that there was essentially no difference in the effects between women and men.

< Table 3 >

Discussion

The hypothesis tested in this study was supported, although unemphatically: results were consistent with an assessment effect, i.e., the group receiving assessment reported lower consumption and fewer alcohol-related problems than did the leaflet only group 12 months after baseline, despite only differing with respect to whether they received a 10-minute web-based assessment of their drinking four weeks after baseline. Notably, however, not all of the comparisons were statistically significant at 12 months and none were significant at 6 months, although all of them were in the hypothesized direction.

There is no obvious reason for the differences between the leaflet only (A) and assessment (B) groups being larger at 12 months than at 6 months. Group A was assessed for the first time at the 6-month follow-up while group B had its second assessment. It is possible that two assessments several months apart were required to motivate changes in drinking over the subsequent 6-month period.

The analytical approach used was conservative, with all analyses planned in advance. Outcomes 1-6 were analysed using procedures which were identical to those described in the pilot trial [6], minimising the risk of Type I error which can result from post hoc comparisons. Outcome 7, AUDIT score at 12 months, was included because the longer follow-up permitted the use of this instrument, most of whose items have a 12-month

reference period. Its inclusion created an opportunity to directly compare 12-month outcomes with baseline data since all the members of both experimental groups completed the AUDIT at baseline.

Importantly, the number of patients who were lost to follow-up did not differ significantly by experimental group, and the overall retention of participants over 12 months was high, such that the group comparisons are unlikely to be biased by differential attrition. The blinding of participants and research assistants to group allocation minimizes the potential for a range of ascertainment biases [2].

The size of the confidence intervals for the outcomes of interest reveal the difficulty in measuring modest effects. Even with initial sample sizes of 146 and 147, confidence intervals spanned 30 percentage points in a mean ratio and two points on the AUDIT. This occurred despite the retention of 86% of participants at the 12-month follow-up assessment. Attrition of less than 20% over 12 months is rare in brief intervention efficacy trials [19].

Theoretical explanations for assessment reactivity are relatively undeveloped. Two competing explanations are presented. The first, *evaluation apprehension*, explains assessment effects in terms of real changes in behaviour produced by a psychological process evoked by assessment [20]. Evaluation apprehension was developed to explain why the presence of other people appears to improve the performance of well-learned tasks and to cause deterioration in the performance of novel complex tasks [21]. On the basis of experimental evidence, Cottrell [20] posited that it is the apprehension in the subject of possible negative evaluation rather than the mere presence of others which

produces these performance effects. Arguably, evaluation apprehension might operate via web-based assessment in persons who know they are to have their drinking evaluated, as was the case in the present trial. Persons exposed to assessment (group B) could be expected to show a larger reduction in their well-learned drinking behaviour than those who were exempted from the four-week assessment (group A). This idea is akin to supposed self-focusing mechanisms in interventions based on motivational interviewing [22], in which recipients of relevant normative feedback are assumed to more closely monitor their own behaviour in relation to norms.

The self-focusing mechanism of feedback may be viewed as a precursor to a self-regulatory mechanism [23], in which individuals integrate comparisons on their current and past alcohol use in relation to what the assessment teaches them about drinking norms. According to this theoretical framework, discrepancy between the individual's self-evaluation and their ideal, produces cognitive dissonance [24], the resolution of which motivates behaviour change. It has been argued [1] that self-regulation processes underpin the FRAMES model used so effectively in motivational enhancement therapy [25], and empirical research on other health behaviour supports the view that feedback plays an important role in self-regulatory processes [1].

An alternative explanation for the results of this paper, namely *social desirability bias* [26], presents apparent assessment effects in terms of measurement artefact, resulting from under-reporting. Social desirability bias occurs when study participants under-report socially stigmatised behaviours or over-report socially desirable ones. For it to explain the observed differences in drinking and alcohol-related problems between the experimental groups in this study, requires that as a consequence of receiving the 10-

minute web-based assessment, patients became more sensitised to under-report their drinking and problems at the 6- and 12-month assessments than did controls. Given the non-judgmental nature of the assessment items and the comfort in which this population group speaks about their drinking [6, 27], it seems unlikely that a social desirability bias was evoked by such a brief assessment, conducted without researchers present, but it cannot be ruled out.

The problem of social desirability bias can be alleviated with the use of bio-markers. We considered and rejected using them because: (1) they are not reliable in this age group, whose drinking is characterised by intermittent bursts of heavy consumption - the bio-markers pick up chronic heavy drinking [28]; and (2) given the brevity of the opportunity for contact at baseline and the web-based follow-up, it would not have been practicable.

This study highlights a challenge in the design and interpretation of treatment studies, and in particular, controlled trials of brief intervention, namely, the need to maximise the experimental contrast while retaining sufficient baseline assessment to permit appropriate statistical control [29]. Alternative approaches include the recruitment of samples of sufficient size to permit assessment of only a portion (a random sample) of each experimental arm at any given follow-up point. The present study utilised the 10-item AUDIT in both experimental groups for the purpose of screening patients into the trial and to increase statistical power in the analysis of outcomes, by using each patient's baseline AUDIT score as a covariate. Accordingly, it could be argued that even the leaflet only group (A) was exposed to some assessment which thereby blurred the experimental contrast. Given the emerging evidence supporting briefer forms of the AUDIT, e.g., the

3-item AUDIT-C [30], there might be benefit in considering their use in clinical trials to help minimise assessment effects.

In summary, the results suggest that brief assessment produced a reduction in hazardous drinking. Controlled trials may, therefore, be systematically under-estimating treatment effects due to blurring of experimental contrast. Implications for practice are that providing routine screening and minimal assessment in primary care may itself be sufficient to produce modest benefits. This may be done electronically, in a way that minimises or eliminates the use of expensive clinician time [31]. While the effect size of this brief assessment was small, the public health significance may be considerable, the value lying in the aggregation of small effects when implemented in large populations.

Implications for research are (1) that in brief intervention trials, careful attention should be paid to the content, length, and frequency of assessment of control participants to avoid Type II error, and (2) efforts should be made to measure and control for social desirability bias.

Acknowledgements

This research was funded by the Alcohol Advisory Council of New Zealand and the Health Research Council of New Zealand. The author are grateful to Associate Professor David Chalmers, and two anonymous reviewers, for helpful comments on earlier versions of the paper.

References

1. Clifford, P.R. and S.A. Maisto, *Subject reactivity effects and alcohol treatment outcome research*. J Stud Alcohol, 2000. **61**(6): p. 787-93.
2. Jadad, A.R., *Randomised controlled trials : a user's guide*. 1998, London: BMJ Books 1998.
3. Mayo, E., *The human problems of an industrial civilization*. 1933, New York: MacMillan.
4. Jonson, H., et al., *Comments on brief intervention of alcohol problems: a review of a review*. Addiction, 1995. **90**(8): p. 1118-21.
5. Bien, T.H., W.R. Miller, and J.S. Tonigan, *Brief intervention: a reply to a review of a review*. Addiction, 1995. **90**: p. 1118-20.
6. Kypri, K., et al., *Web-based screening and brief intervention for hazardous drinking: A double-blind randomised controlled trial*. Addiction, 2004. **99**(11): p. 1410-7.
7. Kypri, K., *Tertiary student hazardous drinking: Epidemiology and development of a brief intervention trial*. 2002, PhD thesis, University of Otago, Dunedin.
8. Scott, E. and P. Anderson, *Randomized controlled trial of general practitioner intervention in women with excessive alcohol consumption*. Drug and Alcohol Review, 1990. **10**: p. 313-321.
9. World Health Organisation Brief Intervention Study Group, *A cross-national trial of brief interventions with heavy drinkers*. American Journal of Public Health, 1996. **86**(7): p. 948-955.
10. Morton, V. and D.J. Torgerson, *Effect of regression to the mean on decision making in health care*. BMJ, 2003. **326**(7398): p. 1083-4.
11. Alcohol Advisory Council of New Zealand, *Alcohol Facts and Effects Leaflet*. 2000, Alcohol Advisory Council of New Zealand: Wellington.
12. Saunders, J.B., et al., *Development of the Alcohol Use Disorders Identification Test (AUDIT): World Health Organisation Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II*. Addiction, 1993. **88**(6): p. 791-804.
13. Kypri, K., S.C.R. Stephenson, and J. Langley, *Episode-centred analysis of drinking to intoxication in university students*. Alcohol & Alcoholism, 2005. **40**(5): p. 447-452.
14. McGee, R. and K. Kypri, *Alcohol-related problems experienced by university students in New Zealand*. Australian and New Zealand Journal of Public Health, 2004. **28**(4): p. 321-323.
15. Kypri, K. and J.D. Langley, *Perceived norms and their relation to university student drinking*. Journal of Studies on Alcohol, 2003. **64**(6): p. 829-834.
16. Dean, C.B., *Overdispersion*, in *Encyclopedia of Biostatistics*, T.C. P. Armitage, Editor. 1998, J. Wiley & Sons: Chichester. p. 3226-3232.
17. Altman, D.G. and J.M. Bland, *Interaction revisited: the difference between two estimates*. Bmj, 2003. **326**(7382): p. 219.
18. Kypri, K., *Electronic screening and brief intervention (e-SBI) for hazardous drinking: A four-arm randomised controlled trial*. Alcoholism Clinical and Experimental Research, 2005. **29**(5 Suppl.): p. 73A

19. Edwards, A.G. and S. Rollnick, *Outcome studies of brief alcohol intervention in general practice: the problem of lost subjects*. *Addiction*, 1997. **92**(12): p. 1699-704.
20. Cottrell, N.B., *Performance in the presence of other human beings: Mere presence, audience and affiliation effects*, in *Experimental social psychology*, C.G. McClintock, Editor. 1968, Allyn & Bacon: Boston.
21. Zajonc, R.B., *Social facilitation*. *Science*, 1965. **149**: p. 269-74.
22. Nye, E.C., G. Agostinelli, and J.E. Smith, *Enhancing alcohol problem recognition: A self-regulation model for the effects of self-focusing and normative information*. *Journal of Studies on Alcohol*, 1999. **60**: p. 685-693.
23. Carver, C.S. and M. Scheier, *On the self-regulation of behavior*. 1998, Cambridge, UK ; New York, NY, USA: Cambridge University Press. xx, 439 p.
24. Draycott, S. and A. Dabbs, *Cognitive dissonance 2: A theoretical grounding of motivational interviewing*. *British Journal of Clinical Psychology*, 1998. **37**(3): p. 355-64.
25. Miller, W.R. and S. Rollnick, *Motivational interviewing : preparing people to change addictive behavior*. 1991, New York: Guilford Press. xvii, 348.
26. Crowne, D.P. and D. Marlowe, *A new scale of social desirability independent of psychopathology*. *Journal of Consulting Psychology*, 1960. **24**: p. 349-354.
27. Kypri, K., *The acceptability of screening and brief intervention to reduce hazardous drinking among university students*. Cutting Edge Conference, Rotorua, New Zealand, 28-30 August, 2000.
28. Babor, T.F., et al., *Talk is cheap: Measuring drinking outcomes in clinical trials*. *Journal of Studies on Alcohol*, 2000. **61**(1): p. 55-63.
29. Vickers, A.J. and D.G. Altman, *Statistics notes: Analysing controlled trials with baseline and follow up measurements*. *British Medical Journal*, 2001. **323**(7321): p. 1123-4.
30. Dawson, D.A., et al., *Effectiveness of the derived Alcohol Use Disorders Identification Test (AUDIT-C) in screening for alcohol use disorders and risk drinking in the US general population*. *Alcohol Clin Exp Res*, 2005. **29**(5): p. 844-54.
31. Kypri, K., et al., *Computerised screening for hazardous drinking in primary care*. *New Zealand Medical Journal*, 2005. **118**(1224): p. 1-9.

Figure 1. Trial schema

Table 1. Gender, age and AUDIT scores of the study groups at baseline

	A		B	
	Leaflet only		Leaflet and	
	group		assessment	
	(n=146)		group	
	(n=147)			
Number (%) females	76	(52)	77	(52)
Mean age (SD)	20.1	(2.2)	20.3	(1.8)
Mean AUDIT score (SD)	15.1	(5.5)	14.9	(5.0)

Table 2. Summary outcome data at 6- and 12-month follow-up assessments

<i>Outcome</i>	<i>Median (range)</i>			
	A Leaflet only group		B Assessment group	
1. Frequency of drinking (number of drinking days in last 2 weeks)				
6 months	4	(0 to 11)	3	(0 to 14)
12 months	4	(0 to 14)	4	(0 to 14)
2. Typical occasion quantity (number of drinks per typical drinking occasion in last 4 weeks)				
6 months	8	(0 to 25)	8	(0 to 25)
12 months	8.5	(1 to 24)	8	(2 to 25)
3. Total consumption (drinks in last 2 weeks)				
6 months	28.5	(0 to 143)	20	(0 to 145)
12 months	30	(0 to 175)	25	(0 to 168)
4. Frequency of very heavy episodic drinking (number of episodes of >80g for women and 120g for men in last 2 weeks)				
6 months	1	(0 to 8)	1	(0 to 6)
12 months	1	(0 to 8)	0	(0 to 8)
5. Personal, social, sexual, and legal consequences of episodic heavy drinking (number of problems – APS; range 0-14)				
6 months	2	(0 to 12)	2	(0 to 9)
12 months	3	(0 to 11)	2	(0 to 9)
6. Consequences related to academic role expectations (score on AREAS; range 0-15)				
6 months	2	(0 to 14)	1	(0 to 12)
12 months	1	(0 to 10)	1	(0 to 14)
8. AUDIT score				
12 months	14	(2 to 30)	13	(1 to 29)

Table 3. Assessment effects based on random effects models without imputation for missing values

<i>Outcome</i>	Assessment effects		P
	Ratio of geometric means* Assessment group / Leaflet only group	(95% CI)	
1. Frequency of drinking			
6 months	0.90	(0.77 to 1.06)	0.20
12 months	0.95	(0.82 to 1.11)	0.53
2. Typical occasion quantity			
6 months	0.92	(0.81 to 1.05)	0.21
12 months	0.98	(0.86 to 1.11)	0.71
3. Total consumption			
6 months	0.87	(0.71 to 1.05)	0.14
12 months	0.82	(0.68 to 0.98)	0.03
4. Frequency of very heavy episodic drinking			
6 months	0.81	(0.58 to 1.13)	0.21
12 months	0.66	(0.47 to 0.91)	0.01
5. Personal, social, sexual, and legal consequences of episodic heavy drinking			
6 months	0.90	(0.74 to 1.09)	0.29
12 months	0.81	(0.67 to 0.99)	0.04
6. Consequences related to academic role expectations			
6 months	0.78	(0.58 to 1.05)	0.10
12 months**	0.85	(0.62 to 1.16)	0.31
7. AUDIT score***			
12 months	-1.63	(-2.65 to -0.62)	0.00

* The exponent of the arithmetic mean of the log- transformed data.

**49 cases in which the participant was no longer at university were excluded from this analysis.

***Beta coefficient based on linear regression with control for baseline AUDIT score.